

# **EXHIBIT 120**

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 1

IN THE UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION

VIDEOTAPED DEPOSITION OF DAVID ESIOBU  
\* \* \* HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY \* \* \*

Reported by:

Connie Recob, CCR, RMR, CRR  
Washington CCR No. 2631  
Oregon CCR No. 15-0436  
Utah CCR No. 1133171-7801  
Idaho CCR No. SRL-1220

DIGITAL EVIDENCE GROUP  
1730 M. Street, NW, Suite 812  
Washington, D.C. 20036  
(202) 232-0646

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 56

1 post training of Llama models. 10:21

2 BY MR. YOUNG: 10:21

3 Q. So would safety -- so would -- so I want to talk briefly 10:21

4 about your responsibilities with respect to safety. 10:21

5 So you have done safety work with respect to filtering 10:21

6 out nonsafe work content with respect to datasets for 10:21

7 training of Llama models, correct? 10:21

8 A. Yes. 10:21

9 Q. Okay. And would it also be fair that you have also done work 10:21

10 with respect to mitigating IP or copyright leakage with 10:22

11 respect to datasets in the Llama model? 10:22

12 MR. WEINSTEIN: Object to form. 10:22

13 THE WITNESS: Some of the work that I've 10:22

14 done has the effect of mitigating regurgitation of training 10:22

15 data. And, yeah, that's kind of the -- so -- so there -- 10:22

16 there are things that we would do during the course of 10:22

17 preparing data that is generally good for the model but also 10:22

18 has the effect of mitigating regurgitation. 10:22

19 BY MR. YOUNG: 10:22

20 Q. So, for example, things like deduplication? 10:22

21 A. Yes. 10:22

22 Q. Okay. So have you, as a general -- just generally in your 10:22

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 57

1 work related to the Llama models, have you studied the Llama 10:22  
2 models' propensity to regurgitate its training model -- or 10:23  
3 training data? 10:23  
4 A. Yes. My -- my evaluations were measuring the rates at which 10:23  
5 I could find generations that match the continuation for 10:23  
6 certain prompts. 10:23  
7 Q. That matched the training data, correct? 10:23  
8 A. The continuations were drawn from the training data, yeah. 10:23  
9 Q. And you did this for Llama-2? 10:23  
10 A. Yes. 10:23  
11 Q. Did you do this for Llama-3? 10:23  
12 A. Yes. 10:23  
13 Q. Okay. And are you doing -- is it anticipated that you will 10:23  
14 do this for Llama-4? 10:23  
15 A. I -- I had planned to, but we're still very much early in 10:23  
16 the -- in the process. 10:23  
17 Q. Okay. Understood. 10:23  
18 MR. WEINSTEIN: Counsel, let me know when 10:23  
19 it's a good time to take a break. 10:23  
20 MR. YOUNG: Yeah, I was just going to 10:24  
21 suggest once I got done with this line of questioning. 10:24  
22 MR. WEINSTEIN: No problem. 10:24

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

		Page 62
1	data.	10:42
2	Q. Do you know if Meta has done any analysis of whether or not	10:42
3	any material included in LibGen is regurgitated by Llama-3?	10:42
4	A. So as part of the memorization analyses that I've done, I	10:42
5	would have drawn samples from all of the datasets that we	10:42
6	use.	10:42
7	Q. Okay. Have you yourself seen any regurgitation from the	10:42
8	LibGen model -- or from the LibGen dataset from Llama-3?	10:42
9	A. So whatever was -- was accepted from the original LibGen,	10:42
10	whatever made it into the model, I drew samples from -- from	10:42
11	that portion of the training data, and I think there -- there	10:42
12	is, like, a nonzero rate across all the datasets that I	10:43
13	tested.	10:43
14	Q. Okay. So would that be a yes?	10:43
15	A. Yes.	10:43
16	Q. Okay. Thank you.	10:43
17	Now, in the course of -- I believe you called it NSFW	10:43
18	filtering. And that's not safe for work, right?	10:43
19	A. Right.	10:43
20	Q. So in the course of your NSFW filtering work with LibGen	10:43
21	fiction, did you examine the contents in LibGen?	10:43
22	A. Yeah. So to -- just to make sure that the filtering was	10:43

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 72

1	A. So Nikolay Bashlykov would have been the -- the point of	10:58
2	contact for this dataset, and so it was likely him.	10:58
3	Q. So do you have an understanding of why copyright notices and	10:58
4	legal disclaimers would be removed before the -- during the	10:58
5	processing of LibGen?	10:58
6	A. So my concern here was kind of the repetitive nature of these	10:58
7	kinds of notices. So this is sort of like a form of	10:58
8	deduplication to avoid, you know -- to -- to avoid kind of	10:58
9	the model, like, regenerating text that it had seen before.	10:58
10	Q. Because would you agree with the general proposition that if	10:58
11	a model is trained on repetitive text multiple times, it is	10:58
12	more likely to regurgitate that -- that portion of text?	10:58
13	A. I would say yes. The tests that I run suggests that	10:58
14	there's -- there's, like, a strong correlation between	10:59
15	repetition and -- in the training data and the rate at which	10:59
16	that will be reproduced.	10:59
17	Q. And that includes repetition across different files, not just	10:59
18	within the same file, correct?	10:59
19	A. Right.	10:59
20	Q. Okay. So, for example -- and would this also be related to	10:59
21	the number of epochs a certain dataset would be run?	10:59
22	MR. WEINSTEIN: Object to form.	10:59

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 83

1 nonzero rate of regurgitation, correct? 11:12

2 A. It's extremely low. I would say -- yeah, I don't -- I don't 11:12

3 remember the exact rates, but it's -- it's very close to zero 11:12

4 for -- for the fine tuned ones. 11:12

5 Q. Okay. Do you know if sci-tech or sci-mag contain academic 11:12

6 textbooks? 11:12

7 A. I -- I kind of like assumed that that kind of thing would be 11:12

8 there, but I haven't looked -- looked into that dataset in 11:12

9 any detail. 11:12

10 Q. Okay. And because you didn't do any work related to sci-mag 11:12

11 or sci-tech? 11:12

12 A. No. 11:12

13 Q. Do you know how Meta obtained copies of LibGen? 11:12

14 MR. WEINSTEIN: Object to form. 11:13

15 THE WITNESS: Not in detail. That 11:13

16 wasn't -- that wasn't a dataset that I worked on in -- in the 11:13

17 kind of like -- I -- I wasn't sort of like -- so each -- each 11:13

18 dataset has, like, a person who's directly responsible for 11:13

19 it. I wasn't directly responsible for that. 11:13

20 BY MR. YOUNG: 11:13

21 Q. Okay. So I understand you weren't responsible for it, but do 11:13

22 you have an understanding of how the dataset was obtained? 11:13

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 101

1 Q. Okay. And then can you read for me -- there's a reference to 11:45  
2 you in the next bullet point, correct? 11:45  
3 A. Yes. 11:45  
4 Q. Okay. Can you please read that next bullet point? 11:45  
5 A. So this is a response to the question. It says, "Yes, we 11:45  
6 will follow a memorization measurement engine approach to 11:45  
7 measure memorization of a 750B model of the IP sensitive 11:45  
8 data" in parentheses "(LibGen synching with David Esiobu)." 11:45  
9 Q. Okay. So do you see the reference to LibGen's IP sensitive 11:45  
10 data? 11:45  
11 A. Yes. 11:45  
12 Q. Do you know why Mr. Bashlykov referred to LibGen as IP 11:45  
13 sensitive data? 11:45  
14 MR. WEINSTEIN: Exclude from your -- 11:45  
15 exclude from your answer any communications you may have 11:45  
16 received from counsel, if applicable. 11:45  
17 THE WITNESS: I don't know exactly how 11:45  
18 Nikolay was thinking about this dataset. But I would 11:45  
19 imagine -- I think it was -- it's widely known that it 11:46  
20 contained books. 11:46  
21 BY MR. YOUNG: 11:46  
22 Q. Okay. And not just books, copyrighted books, right? 11:46

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 102

1 A. Yes. 11:46

2 Q. Thank you. 11:46

3 Now, did you conduct memorization evaluations of the 11:46

4 LibGen model -- or the LibGen dataset with the model 11:46

5 consistent with what was described here? 11:46

6 A. So I ended up not running the memorization evaluations for -- 11:46

7 for these experiments because I had some other projects I had 11:46

8 going on at the time, and so we didn't actually close on -- 11:46

9 on these experiments. 11:46

10 Q. Okay. Do you know if anyone actually conducted the 11:46

11 memorization evaluation experiments for LibGen? 11:46

12 A. If anyone followed up on these experiments in terms of the 11:46

13 impact on memorization, I -- I'm not aware of anyone doing 11:46

14 that. 11:46

15 Q. Okay. So to your -- just to make sure I -- I understand, to 11:46

16 your knowledge, you're not aware of anyone at Meta conducting 11:47

17 the experiments to determine the affect of LibGen on 11:47

18 memorization, correct? 11:47

19 MR. WEINSTEIN: Object to form. 11:47

20 THE WITNESS: I'm not aware of anyone 11:47

21 evaluating the checkpoints that -- that Nikolay 11:47

22 trained described here. The -- the tests I've done for 11:47

12/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

David Esiobu

Page 271

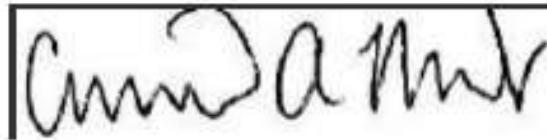
1

## REPORTER'S CERTIFICATE

2

3 I, CONNIE A. RECOB, the undersigned Certified Court  
4 Reporter, authorized to administer oaths and affirmations in  
5 and for the States of Washington, Oregon, Utah and Idaho, do  
6 hereby certify that the sworn testimony and/or proceedings, a  
7 transcript of which is attached, was given before me at the  
8 time and place stated therein; that any and/or all  
9 witness(es) were duly sworn to testify to the truth; that the  
10 sworn testimony and/or proceedings were by me  
11 stenographically recorded and transcribed under my  
12 supervision, to the best of my ability; that the foregoing  
13 transcript contains a full, true, and accurate record of all  
14 the sworn testimony and/or proceedings given and occurring at  
15 the time and place stated in the transcript; that a review of  
16 which was requested; that I am in no way related to any party  
17 to the matter, nor to any counsel, nor do I have any  
18 financial interest in the event of the cause.

19 WITNESS MY HAND and SIGNATURE this 18th day of  
20 December, 2024.



21 /s/CONNIE A. RECOB, RMR, CRR  
22 Washington CCR No. 2631  
Oregon CCR No. 15-0436  
Utah CCR No. 1133171-7801  
Idaho CCR No. SRL-1220